

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/701,278

DATE: 03/11/97

TIME: 15:44:52

INPUT SET: S16092.raw

This Raw Listing contains the General
Information Section and up to the first 5 pages.

#5

ENTERED

SEQUENCE LISTING

(1) General Information:

(i) APPLICANT: Anderson, David J.
Saito, Tetsuichiro

(ii) TITLE OF INVENTION: A NOVEL HOMEODOMAIN PROTEIN

(iii) NUMBER OF SEQUENCES: 22

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
(B) STREET: Four Embarcadero Center, Suite 3400
(C) CITY: San Francisco
(D) STATE: California
(E) COUNTRY: United States
(F) ZIP: 94111

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US 08/701,278
(B) FILING DATE: 22-AUG-1996
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Silva, Robin M.
(B) REGISTRATION NUMBER: 38,304
(C) REFERENCE/DOCKET NUMBER: A-63770-1

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (415) 781-1989
(B) TELEFAX: (415) 398-3249

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2424 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/701,278DATE: 03/11/97
TIME: 15:44:57

INPUT SET: S16092.raw

47 (D) TOPOLOGY: unknown

48

49 (ii) MOLECULE TYPE: DNA (genomic)

50

51

52

53

54 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

55

56 GCAGAGGTAG GCAGGGTTCC CGAGCCGCTC TCCC GGCTCC CTGCTCTGGG CCTTGGGGCT 60

57

58 CCACCGGCTT CTTGGCCCGA GCTGCTGCGC GTGCAGATGG CCTTGCGCGA TCGCCGGACC 120

59

60 CCGCTGCGGT GGCCAAGTGC AGGGCTTGTG GCTGGGACCC CTGAGAACCA GGAGCCAGAC 180

61

62 TGTGCTCAGC TTGCCAGGCC GGAGCCACGC ACGGGCACAA GTCTGTCAGG CCGCCATCAG 240

63

64 TCCTGGTCCA GCCGTCAGGG CCCATCCGAC CGTCGGCGAT GTTTTATTTT CACTGCCCCG 300

65

66 CACAGCTAGA GGGCACAGCG CCTTTTGGTA ACCACTCTAC GGGGGATTTT GATGATGGGT 360

67

68 TTCTTAGAAG AAAACAGCGC AGAAATCGGA CAACCTTCGC TCTTCAGCAG TTGGAAGCTC 420

69

70 TGGAGGCAGT CTTTGCCCCA ACACACTACC CAGATGTCTT CACCAGAGAA GAGCTAGCCA 480

71

72 TGAAAATAAA CCTCACAGAA GCCAGAGTGC AGGTTTGGTT CCAGAACCGA AGAGCCAAGT 540

73

74 GGAGGAAGAC AGAGAGAGGG GCCTCTGACC AGGAACCAGG GGCTAAGGAA CCCATGGCAG 600

75

76 AGGTGACACC ACCCCCAGTG AGGAACATCA ACTCTCCACC CCCAGGGGAC CAGGCCCCGG 660

77

78 GCAAGAAGGA GGCCCTGGAG GCCCAGCAGA GCCTGGGACG CACAGTGGGC CCCGCCGGGC 720

79

80 CTTTCTTCCC CTCCTGCTTG CCAGGGACCC TCCTGAACAC AGCCACTTAT GCCCAGGCCC 780

81

82 TGTCCCATGT GGCATCTCTG AAAGGGGGCC CACTGTGCTC TTGCTGCGTC CCAGACCCTA 840

83

84 TGGGGCTCTC CTTCCCTCCC ACTTACGGTT GCCAGAGTAA CCGCACAGCC AGCGTGGCTG 900

85

86 CCCTGCGCAT GAAGGCCCGC GAGCATTCAG AAGCGGTCCT GCAGTCTGCC AACCTTCTGC 960

87

88 CGTCCACCAG CAGCAGCCCC GGCCCTGCCT CCAAGCAGGT GCCTCCAGAA GGCAGCCAGG 1020

89

90 ACAAGCCCTC CCCAACGAAG GAACAGAGCG AGGGAGAGAA GAGCGTATGA GGGTCCGGAG 1080

91

92 AACCAGCTG GGAGCCCTGC CCACCCCTGC TTCTCTCAGC CTCAGCCCTG CCAGCCTCTG 1140

93

94 AACCACAAGG AGTAGCCACC TCCTCATGGA TCTGACAGGG CAAACGGGAC CTGCAAGCTG 1200

95

96 GTTGAGACCT GAAGAGTCCC TCTAGAATTC TGCTGGTAGG CTGTGTTGTT CTCGCTTTTC 1260

97

98 CTTTGGTGAC ATTTTCCGAT GGCTCTTAGT GACTCTGGAC ACTGCTCTGT GATGAGGTCC 1320

99

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/701,278

DATE: 03/11/97

TIME: 15:45:02

INPUT SET: S16092.raw

100	CTGTTTTTTTG CTTTTTGTTT TGTCTCTTTT TTTTTGTTTT GTTTTGTTTT ATTTTCCAGG	1380
101		
102	CCAAGCAGCC TTGGAGCAAA GCAGATTAGT TTATTCCACC ATCCTTCTTG AGATATCTGG	1440
103		
104	GAAGGTCTTG TCAATTCCAA GGACTGTGGC AAGGATCATC CGTGAAAGAT GCCAAGAAGT	1500
105		
106	GACATCTCAT GACAGGAAAT GAGACGGGCA CTCCCATATT GCTTAAGAAC CACAGAACTG	1560
107		
108	GTGGACTATC AGCCAGTTCT CACTCCCTTC AGCCAGGACT GGCATCGGCC TCCTTTGTCT	1620
109		
110	TGTTTAAAGG AATTAGCTGA GGTTTTGGCT AGGAAGTGAC AAGATATGGG CTGAAGACAT	1680
111		
112	TGTGGTCCTG ACCCTAGCAG ATCTCCCTGG GCACATCTGA CCTGGTCCAG TCAGGCAGGT	1740
113		
114	TGTCAGTTTG GGGATGGGGG CTGCTCTGCT GATTCTGTGT GTGGGTTCCT TGCAATTAGA	1800
115		
116	GTGTTCACTT GCAGGCCCCG CTCTCTTCAG AAGAGTGATG GGAAGTTCAC CAATCAGAAT	1860
117		
118	GTAGCTTTGT AGCCCAGGAA AGGACCAGAG TCCTTGAAGC GGTAGGAAAT CCCTAGGAAG	1920
119		
120	GCCCCTTAAA TACTTATGCC CAGATGAGCT GCCCTTCTTC CTATCCCCGT ATGTCGAGAG	1980
121		
122	GTTGACGAGA CAGGAAAGCC AGGAAGATGA CTCCGTGTGG CAGAAGAGAA TGGAGTCCAA	2040
123		
124	AGGGCCAAC TTTACAGAGA TTTCTGCCGC AGTTTAGCGT GGCTGTGTTC TTTACGCGA	2100
125		
126	TGGTGACTTC GGAGAGATCA GAGGGAGATG TGCAATAGCA TGAGCCCCGC TCCTGGCCCCG	2160
127		
128	GGTCCTGGAA AGGTTGTGGT TGTTTGGTGG CTTTGGCTGA TGATGTTTCC ACGCAAACAG	2220
129		
130	ATATTGCTTT CATGATGGCT GTTCTCATTT CAGTTCTGAT AATCGAGACG CTGTGCTCCC	2280
131		
132	AGGCGCTCTG CCTCCCCCTA ACTCTTCAGG AGCACCCCCT CCCCTGTAAT ACTCCTAAGT	2340
133		
134	GTATCGTGCC TCACTTACGG TTACTGCAAC ACATTTGATG GAACACACTG TCTCCTTTAA	2400
135		
136	AAAAGAAAAA AAAAAAAAAA AAAA	2424

137

138 (2) INFORMATION FOR SEQ ID NO:2:

139

140 (i) SEQUENCE CHARACTERISTICS:

141 (A) LENGTH: 263 amino acids

142 (B) TYPE: amino acid

143 (C) STRANDEDNESS: unknown

144 (D) TOPOLOGY: unknown

145

146 (ii) MOLECULE TYPE: protein

147

148

149

150

151 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

152

RAW SEQUENCE LISTING PATENT APPLICATION US/08/701,278

DATE: 03/11/97

TIME: 15:45:07

INPUT SET: S16092.raw

153	Met	Phe	Tyr	Phe	His	Cys	Pro	Pro	Gln	Leu	Glu	Gly	Thr	Ala	Pro	Phe
154	1				5					10					15	
155																
156	Gly	Asn	His	Ser	Thr	Gly	Asp	Phe	Asp	Asp	Gly	Phe	Leu	Arg	Arg	Lys
157				20					25					30		
158																
159	Gln	Arg	Arg	Asn	Arg	Thr	Thr	Phe	Ala	Leu	Gln	Gln	Leu	Glu	Ala	Leu
160			35					40					45			
161																
162	Glu	Ala	Val	Phe	Ala	Gln	Thr	His	Tyr	Pro	Asp	Val	Phe	Thr	Arg	Glu
163		50					55					60				
164																
165	Glu	Leu	Ala	Met	Lys	Ile	Asn	Leu	Thr	Glu	Ala	Arg	Val	Gln	Val	Trp
166	65					70					75				80	
167																
168	Phe	Gln	Asn	Arg	Arg	Ala	Lys	Trp	Arg	Lys	Thr	Glu	Arg	Gly	Ala	Ser
169					85					90					95	
170																
171	Asp	Gln	Glu	Pro	Gly	Ala	Lys	Glu	Pro	Met	Ala	Glu	Val	Thr	Pro	Pro
172				100					105					110		
173																
174	Pro	Val	Arg	Asn	Ile	Asn	Ser	Pro	Pro	Pro	Gly	Asp	Gln	Ala	Arg	Gly
175			115					120					125			
176																
177	Lys	Lys	Glu	Ala	Leu	Glu	Ala	Gln	Gln	Ser	Leu	Gly	Arg	Thr	Val	Gly
178		130					135					140				
179																
180	Pro	Ala	Gly	Pro	Phe	Phe	Pro	Ser	Cys	Leu	Pro	Gly	Thr	Leu	Leu	Asn
181	145					150					155					160
182																
183	Thr	Ala	Thr	Tyr	Ala	Gln	Ala	Leu	Ser	His	Val	Ala	Ser	Leu	Lys	Gly
184					165					170					175	
185																
186	Gly	Pro	Leu	Cys	Ser	Cys	Cys	Val	Pro	Asp	Pro	Met	Gly	Leu	Ser	Phe
187				180					185					190		
188																
189	Leu	Pro	Thr	Tyr	Gly	Cys	Gln	Ser	Asn	Arg	Thr	Ala	Ser	Val	Ala	Ala
190			195					200					205			
191																
192	Leu	Arg	Met	Lys	Ala	Arg	Glu	His	Ser	Glu	Ala	Val	Leu	Gln	Ser	Ala
193		210					215					220				
194																
195	Asn	Leu	Leu	Pro	Ser	Thr	Ser	Ser	Ser	Pro	Gly	Pro	Ala	Ser	Lys	Gln
196	225					230					235					240
197																
198	Val	Pro	Pro	Glu	Gly	Ser	Gln	Asp	Lys	Pro	Ser	Pro	Thr	Lys	Glu	Gln
199					245					250					255	
200																
201	Ser	Glu	Gly	Glu	Lys	Ser	Val									
202					260											
203																

(2) INFORMATION FOR SEQ ID NO:3:

205

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/701,278

DATE: 03/11/97

TIME: 15:45:12

INPUT SET: S16092.raw

206 (i) SEQUENCE CHARACTERISTICS:
207 (A) LENGTH: 70 amino acids
208 (B) TYPE: amino acid
209 (C) STRANDEDNESS: unknown
210 (D) TOPOLOGY: unknown
211
212 (ii) MOLECULE TYPE: protein
213
214
215
216
217 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:
218
219 Gly Phe Leu Arg Arg Lys Gln Arg Arg Asn Arg Thr Thr Phe Ala Leu
220 1 5 10 15
221
222 Gln Gln Leu Glu Ala Leu Glu Ala Val Phe Ala Gln Thr His Tyr Pro
223 20 25 30
224
225 Asp Val Phe Thr Arg Glu Glu Leu Ala Met Lys Ile Asn Leu Thr Glu
226 35 40 45
227
228 Ala Arg Val Gln Val Trp Phe Gln Asn Arg Arg Ala Lys Trp Arg Lys
229 50 55 60
230
231 Thr Glu Arg Gly Ala Ser
232 65 70
233
234 (2) INFORMATION FOR SEQ ID NO:4:
235
236 (i) SEQUENCE CHARACTERISTICS:
237 (A) LENGTH: 70 amino acids
238 (B) TYPE: amino acid
239 (C) STRANDEDNESS: unknown
240 (D) TOPOLOGY: unknown
241
242 (ii) MOLECULE TYPE: protein
243
244
245
246
247 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:
248
249 Leu His Glu Lys Arg Lys Gln Arg Arg Ile Arg Thr Thr Phe Thr Ser
250 1 5 10 15
251
252 Ala Gln Leu Lys Glu Leu Glu Arg Val Phe Ala Glu Thr His Tyr Pro
253 20 25 30
254
255 Asp Ile Tyr Thr Arg Glu Glu Leu Ala Leu Lys Ile Asp Leu Thr Glu
256 35 40 45
257
258 Ala Arg Val Gln Val Trp Phe Gln Asn Arg Arg Ala Lys Phe Arg Lys

PAGE: 1

SEQUENCE VERIFICATION REPORT
PATENT APPLICATION US/08/701,278

DATE: 03/11/97
TIME: 15:45:19

INPUT SET: S16092.raw

Line

Error

Original Text